

クレジット:

UTokyo Online Education Education コンピュータシステム概論 2018 小林克志

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



Python の基本

この資料は [The Python Tutorial \(https://docs.python.org/3.6/tutorial/index.html#the-python-tutorial\)](https://docs.python.org/3.6/tutorial/index.html#the-python-tutorial) (日本語版 (<https://docs.python.jp/3/tutorial/>)) および [Python for Data Analysis:Wrangling with Pandas, Numpy and IPython \(http://shop.oreilly.com/product/0636920050896.do\)](http://shop.oreilly.com/product/0636920050896.do)を参考に作成した。

このチュートリアルでは他のプログラミング言語を習得している方を想定し、Python の特徴などを説明する。

システムコール(参考)

データ処理ではより高度な機能をもつ `pandas` のファイル I/O を利用する方が良い。

システムコールはプログラムがオペレーティングシステムを介してコンピュータ資源、例えばファイルシステムやネットワーク、へのアクセス手段を提供する。

ファイル I/O

ここでは、ファイルシステムを例に紹介する。もっとも基本的な書き方は以下のとおり。このとき、一度開いたファイル構造体 `.close()` を忘れないこと:

```
In [20]: path = "/usr/share/dict/words"

f = open(path)

i = 0
for line in f:
    i = i + 1
    print(line.strip())
    if i >= 10:
        break

print()
print(f.closed)           # ちゃんとクローズされていない
f.close()
print(f.closed)
```

```
A
a
aa
aal
aalii
aam
Aani
aardvark
aardwolf
Aaron

False
True
```

関数にもよるが with を利用すればコードブロック終了時、インデントが戻された時点で、必要な後処理を自動的におこなう。open() の場合は close() 処理がおこなわれる:

```
In [19]: path = "/usr/share/dict/words"
```

```
with open(path) as f:
    i = 0
    for line in f:
        i = i + 1
        print(line.strip())
        if i >= 10:
            break

print()
print(f.closed)
# f.close() ここでは不要
```

```
A
a
aa
aal
aalii
aam
Aani
aardvark
aardwolf
Aaron
True
```

lines に全てのデータを読み込む場合は内包表記リストで、個別に読み込む場合は generator で、それぞれ以下のように記述してもよい:

```
In [1]: path = "words"
lines = [x.rstrip() for x in open(path)]
lines[0:10]
```

```
Out[1]: ['A',
         'a',
         'aa',
         'aal',
         'aalii',
         'aam',
         'Aani',
         'aardvark',
         'aardwolf',
         'Aaron']
```

```
In [2]: path = "words"
lines = (x.rstrip() for x in open(path))
i = 0
for line in lines:
    i = i + 1
    print(line)
    if i > 9:
        break
```

```
A
a
aa
aal
aalii
aam
Aani
aardvark
aardwolf
Aaron
```